DOCUMENT RESUME

ED 109 255 TM 004 720

AUTHOR Shepard, Lorrie
TITLE Reporting the Results of Statewide Assessment.
PUB DATE [Apr 75]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D. C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Educational Assessment; Guidelines; *Information Dissemination; Program Descriptions; Program Planning; *State Programs; *Test Results

ABSTRACT

Reporting the results of statewide assessment looms as a problem as more states pass from the planning to implementation phase in their assessment programs. When energies are focused on the purpose of the assessment, formulating objectives, and instrument construction, reporting takes a back seat because it happens last. There are some general principles to be followed in order to report effectively the results of a large scale assessment program. This paper begins with several recent references on how to report the results of large scale assessment programs. The remainder of this paper is intended to provide specific new thoughts for implementation of old principles. The ultimate success of state assessment programs will depend on how well assessment results are reported to their various audiences. In this paper, the most compelling recommendations for improving reporting practices are plan ahead, develop different reports for different audiences, and field test report formats to determine the language and content that are most meaningful to respective audiences. Reporting should receive the same careful attention as instrument construction with sufficient opportunity for feedback from intended users. (Author/DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED109255

TM004 720

REPORTING THE RESULTS OF STATEWIDE ASSESSMENT

Lorrie Shepard
University of Colorado

Reporting the results of statewide assessment looms as a bigger problem as more states pass from the planning to implementation phase in their assessment programs. When energies were focused on the purpose of the assessment, formulating objectives, and instrument construction, reporting took a back seat because it would happen last. That was, of course, our first mistake, which is one of the points made in this paper. Reporting is a bigger problem now because (1) now we actually have to do it, (2) we haven't given reporting the same kind of attention we've given to problems in test development or sampling, and (3) all the errors in other aspects of the assessment program accumulate in the reports.

There are some general principles to be followed in order to report effectively the results of a large scale assessment program. These principles are not new. For example, Bob Stake suggested some time ago that evaluation reports should be tailored for specific audiences. If state agencies and contractors are having trouble writing reports that will be read, it is not because they haven't heard the rules often enough, but because there is much that can go wrong between the theory of what makes a good report and putting it into practice. This paper begins with several recent references on how to report the results of large scale assessment programs. The remainder of the paper is intended to provide specific new thoughts for implementation of old principles.

The single most important reference on assessment is Frank Womer's monograph, Developing a Large Scale Assessment Program. It is commendable not just because it provides useful guidelines, but also because the author suggests specific reporting strategies within the context of a total assessment plan. The Cooperative Accountability Project also sponsored a three-part document, A

Dissemination System for State Accountability Programs, written by professors of communication, Bettinghaus and Miller. One of its most valuable elements is a set of recommendations for dealing with the news media. A final reference is Ed Larsen's Suggestions for Talking to School-Community Groups about Testing and Test Results. Additional references should include the proceedings of numerous conferences for state assessment personnel where reporting problems and solutions have been major topics, such as the ETS Conference for directors of state testing programs, the National Assessment workshops and a meeting of eight states in Florida sponsored by USOE. Unfortunately, this wisdom has not been published. The attempt in this paper to build upon the knowledge already available is therefore limited by the author's attendance at some but not all of the meetings.

1. Plan Ahead: Reporting should receive as much attention as test construction.

The cardinal rule for good reporting is to decide what should be reported before planning the assessment. It should be the first step, not the last. Only the minor cosmetic aspects of reporting can safely be postponed.

2. Different Reports for Different Audiences.

A second overriding principle, whose elaboration will provide for implementation of the first, is that reports should be tailored for specific audiences. Different audiences need different information. This has implication for both the content and format of assessment reports.

Choosing the appropriate content is clearly the more important consideration. How to properly display irrelevant data is a foolish question. The content of a report not only determines whether it will be useful to its audience; the intended content of reports will dictate the design of the assessment, instrumentation, data collection, and analysis. Further discussion of assessment content has been omitted, however, because it receives full attention in the accompanying paper by Jim Impara.

In this paper, the more mundane questions of format are addressed. What is the proper length, wording, organization and medium that will suit each audience? These may in fact be the variables that determine whether the well-chosen information is received.

The best technique for identifying audiences is that of sample reports, recommended by Frank Womer and earlier by Bob Stake. Sample reports can be simple sentences that exemplify the choices such as whether the results are to be reported for a state, district, school, or individual, and whether the criteria will be percent passing a pre-specified number of items or in reference to some norm. Frank Womer offers ten of these examples as part of his discussion about how to determine the purpose of a large scale assessment. These decisions must be made at the outset, and according to Womer, the "best way to 'force' those persons who establish policy and purpose is to give them a series of possible types of reports and have them decide which ones provide the type of information they really want." (p. 19)

Once audiences have been identified as well as the information appropriate to each, then the selected sample reports become the basis for further elaboration of specific reporting strategies.

Planning sessions could begin by making two lists, one of the possible audiences of the assessment and the other of types of information available. Obviously, the more fully implemented the assessment program is by the time the planning meeting takes place, the more constraints there will be on the second list. A matrix can then be constructed with potential audiences along one dimension such as legislators, classroom teachers, parents, superintendents, reading specialists and educational researchers, and along the other dimension would be examples of assessment results such as a pupil's scores on reading objectives, statewide averages in comparison to national norms, or district level

subtest scores in relation to pre-specified criteria. The matrix is useful initially to make certain that none of the important types of information or audiences are left out. Real progress will be made in planning, however, when the matrix begins to collapse; once it is possible to identify similarities in the needs of certain audiences, then it is possible to specify a manageable number of report types with each addressed to specific information requirements.

Planning for reporting is complicated somewhat by the interaction between report content and a third dimension, report format. Clearly content choices should govern the selection of report format. But it is only possible to present certain kinds of information in a one-minute television segment. There will be some audience characteristics, such as technical understanding, political perspective, or attention span, that will make some modes of presentation unacceptable. These limitations on format may in turn cause reconsideration of the kinds of information that can reasonably be included in each basic report type.

3. All Reports Aren't Written Reports.

When speaking of state assessment reports, one usually means written reports. This habit occurs because written reports are proportionally the biggest share of state department reports, certainly if measured by weight. Many of the suggestions offered in this paper are most appropriate for written documents, but this should not cause us to overlook the virtues of other media. Slide shows or filmstrip presentations are visually alluring and frequently hold an audience's attention longer than the same words and graphs would in a written document. There are costs involved, of course, and a slide presentation usually requires more in dollars and in staff expertise. Written reports have more to recommend them than lower development costs. They are also more easily referenced; two months later, it is easier to refer back to page 27 than to return to the third segment in a film presentation. In addition, the total cost of reporting will

involve a trade-off between the cost of each copy and the number of individuals or agencies who can have their own copy of the report. When enormous numbers are required, or when each report is individualized (e.g., different information for each district) then audio-visual presentations are less feasible. Nevertheless, for one-time audiences who need an overview of the state results, a media presentation may be the most effective.

4. Personal Contact Enhances Reporting.

The examples of different media offered above are all potentially long-distance transmittors of assessment results. But, an important rule for successful reporting is that the reports should be delivered personally. If those who are responsible for reporting results can convey them directly to the respective audiences, whether state legislators or district superintendents, there is an increased likelihood that the message will be received. Face-to-face contact ensures that the reports will be looked at and provides the opportunity to answer technical or interpretive questions that cannot be answered by a written document.

This may seem an outrageous proposal, especially from a Californian who should know that such a practice is impossible. But even in California, where an assessment staff of six faces more than 1,000 districts and 5,000 schools, some personal contact is provided by means of workshops. Area workshops are held where district personnel can receive the reports and suggestions about how to read and interpret the reports. The district personnel then provide a direct contact for school principals and teachers. In most states, state assessment staff make verbal reports to the legislature, but in many states this is the only face-to-face reporting. Pennsylvania's example is rare where state assessment staff visit every district involved in the assessment. Some additional states hold public meetings throughout their states to provide a forum for the

dissemination of assessment results. Such meetings or the kinds of workshops described in California should be considered as a minimal response to the requirement for personal contact in reporting results.

5. Reports Should be Journalistic Rather Than Scholarly.

Reports should be less like dissertations and more like newspapers. This admonition was prompted by the dreariness of many written reports, but it has implication as well for the organization of information in slide shows and verbal presentations.

Dissertations are typically organized following a laborious logic: albeit based on the reasoning of the scientific method, such organization is useful only if one reads the entire document. Similarly, state assessment reports usually begin at the beginning with the purposes of the assessment, goal development, the hierarchy of performance objectives, etc., leaving the important information, the results, buried in the middle between the introduction and appendices.

The recommendation for a more journalistic style is based on the assumption that most readers will not read the entire report even if, guided by principle number two, its length and content have been fashioned especially for them. Following the old who, what, when, where, and how paradigm, information should be organized so that the most salient results are presented first. Each succeeding paragraph that is added to the narrative should summarize the most important information from what remains unsaid. This ordering of information from most to least newsworthy will ensure the greatest information pay-off for each reader regardless of where he stops reading. Of course, when the assessment results are extensive, with separate categorization by region, district, or background variables, each reader will have a different "most important part." But there are still some general rules to follow. At least make the results easy to find and separate from the purposes and procedures of the assessment. At

least give the "big group" results or statewide findings before reporting for subgroups or by background variables.

Newspaper reporters should also be mimicked when selecting those aspects of the assessment results which are most important to respective audiences. In general, differences are more interesting than similarities. If the results this year are the same as last year's in all subject areas except math, then the headline information is whether the math scores are up or down.

6. Reports Should be Shorter.

Principle number six follows directly from principle number five. In addition to saying the most important information first, sometimes reports should then stop immediately. The length of a report is one of the variables that can be most successfully manipulated to alter reports for their respective audiences. National Assessment uses "Executive Summaries" and some states are beginning to create separate abbreviated documents especially for the legislative audience. This practice should be expanded. If assessment results are going to attract the attention of legislators, taxpayers, parents, and newsmen, they will have to be brief. Why not develop a single page summary entitled "Major Findings of the 1975 Statewide Assessment?" Such a page should be available separately as well as being the first page of thicker documents. Those who are intrigued by the information in the single page can certainly, and are more likely to, read further. Save the elaborate breakdowns by subject subcategories, such as prefixes and suffixes or consonant blends and digraphs, for the subject matter experts.

7. Use Data Displays Wisely and They will Carry the Report.

Graphs and tables are frequently more effective summaries of results than pages and pages of narrative. The first rule for using them wisely is that too many will spoil their effectiveness. Reams of tables should be relegated to the

appendix leaving only selective examples in the text of the report. This rule applies as well to slide shows and film presentations.

Tables should be well labeled so that they stand on their own. A second and seemingly contradictory recommendation is that the use of tables should be coordinated with the text so as to provide the reader with examples of how to interpret the particular intersections of rows and columns. The possible contradiction is based on the assumption that the author of the report cannot know how each reader will extract information from the report. Some readers attend to only the tables and graphs, others read the words and skip the figures.

Reports are frequently computer generated, especially when results are reported for individual districts and schools. In these instances, pre-printed forms or computer written sentences should be used to clarify the meaning of numbers in the giant tables that result. In California, one of the best received forms was the report to districts and schools of the 1974 second and third grade reading assessment results. The two page computer generated form had four major sections: total reading test results in comparison to national norms, total test results in comparison to predicted scores based on multiple regression, a summary of background data, and results by subcontent areas in reading (e.g., phonetic analysis, consonants and vowels, under word identification skills). The first section is reproduced here:

MEAN TEST SCORE

The Mean *Reading Test* Scores for second and third grade pupils in your district and school are shown in the last two columns. The means are also given for all second and third grade pupils in the State of California and for the Nation (as derived from Publishers' Norms).

The Mean Test Score is the average percent of questions on the total *Reading Test* that were answered correctly in your district or school. You may want to think of this score as the percent of questions on the total test that the "average child" answered correctly.

THE AVERAGE SECOND GRADER IN YOUR DISTRICT ANSWERED 76.5 PERCENT OF THE QUESTIONS CORRECTLY. IN CALIFORNIA THE AVERAGE SECOND GRADER ANSWERED 66.6 PERCENT OF THE QUESTIONS CORRECTLY.

Reading Test Results

Grade Level	Mean Test Score (Average Percent of Questions Answered Correctly)			
	National	State	District	School
Grade 2	63.9	66.6	76.5	
Grade 3	79.7	80.2	89.3	

The computer written sentences are redundant with what is said in the table and in the general description. But it was the individualized sentences that were the most popular aspects of the report. They made it possible for those who were unfamiliar with the technical terms to verbalize the assessment results.

Graphs are often better than tables because they make similarities and differences more visible. Graphs may be the only "analysis" that is required for some audiences. The best rules for using graphs wisely are the old rules. For example, quantitative scales should include zero so as not to exaggerate small differences. A good reference is Chapter Three in Statistical Methods in Education and Psychology by Glass and Stanley (1970). In addition, the successfulness of graphs will be enhanced by good labelling and by one-liners that repeat the message conveyed by the graph. In the California example used above, the following sentence appeared under a graph where the individual district score was displayed in relation to the distribution of district scores statewide: "When district averages are ranked statewide, the middle score (median) is 67.37 for grade 2 and 81.85 for grade 3. These scores can be thought of as the performance of an 'average district' in the state." Further information was then provided regarding the relationship of a district's score to the percentile scale.

8. Save Technical Explanations for Footnotes or Technical Supplements.

Statistical analyses are pursued presumably to increase the meaningfulness of raw score results. They should never obscure the information. Unfortunately, when a statistical tool is used, it sometimes becomes so dominant in the narrative that it overshadows the assessment results. For example, if "the standard error of the difference" is repeated a dozen times in a single paragraph, the reader is likely to worry more about his understanding of this statistic rather than the implications of the assessment results. It is preferable to translate

statistical principles into general rules of thumb for the reader so that the statistical qualifiers will not have to be repeated at every turn. Here is another example from the California report intended for district and school personnel. Speaking of 14 subscores from the reading test:

> The percentile ranks are reported in the fourth column as bands or ranges rather than single points in order to show the error associated with test scores. The bands are shown on the graph to discourage over-interpretation of small differences in subtest scores. The more important differences are likely to be those where the bands do not overlap.

In the same report, multiple regression was used to compute expected scores for each school or district for comparative purposes. In the brief narrative on the computer form, the term multiple regression was not used; instead, an effort was made to present the same concept in lay terms:

> The numbers in the second column denote the range of scores that are most likely to be obtained by districts or schools like yours. This "Comparison Score Range" was computed using the background factors below.
>
> IN GRADE 2 MOST SCHOOLS LIKE YOURS SCORED IN THE RANGE FROM THE 41ST TO THE 69TH PERCENTILE.

Technical supplements were provided for those who were interested in how each of the background variables was operationalized and in the beta weights used in the regression equation.

9. Overcome Statistical Conservatism.

Even more bothersome than the technical vocabulary of the statistician is the conservative training from inferential statistics which says, "we can never know anything with certainty." On occasion, this conservatism has the effect of making all assessment results seem equivocal. This leaves non-technical audiences with the feeling that none of the results can be trusted and wondering what good the assessment was anyway.

Of course, there are errors due to sampling and imperfections in the measurement device, and less estimable errors such as those due to differences in test administration or school by test-content interactions. But presumably, these are not so enormous as to entirely invalidate the assessment results. If this were the case, the assessment program should have been called off short of printing the results. The appropriate statistical tools should be used to estimate the magnitude of errors associated with particular scores or differences in scores. These should be used to preclude over-interpretation of small differences. In addition, some initial disclaimer may be called for concerning the appropriateness of the test content for making some decisions but not others. But then, the disclaimers and equivocations should cease. The report should make the best statement possible using the decision rules from inferential statistics, and then let the statement stand. If sixth grade math scores have gone up more than could be accounted for by sampling fluctuation or differences in test administration, then the report writer should say that there has been a trustworthy change in the level of pupil performance.

10. Don't Obscure the Information.

In a few noteworthy instances, assessment results have been meaningless to their audiences not because of statistical conservatism or lack of journalistic skill, but because it has been the intention of the report authors or the state agency to obscure the information. Repeated examples of misinterpretation of assessment results by newspapers and legislators have made state assessment personnel protective of themselves and local educators. It is obviously very difficult to get an undistorted message delivered to the public sector. But it is unforgivable to react by shrouding the assessment results in statistical or educational jargon. The most common ploy is to produce such a surfeit of data

that neither the press nor anyone else can make any sense out of it. Another example occurred at a recent workshop where representatives from one state bragged about their practice of reporting district averages in relation to the state average for pupils rather than in comparison to the mean or median of district scores. Because large districts in their state tend to be low scoring districts, the result was that more than half of the districts had results which could be called "above average." This they believed was the best of all possible worlds. These comments are not intended to dispute the comparative strategy used; each reference statistic has different meaning and selection should be based on the purpose of the comparison. What was clearly wrong, in the opinion of this author, was the boast that significant audiences were fooled rather than enlightened by the assessment results. How can such practices be sanctioned by the same individuals who advocate assessment because it will provide useful information?

11. Make Comparative and Interpretive Information a Part of the Report.

Comparisons are essential if respective audiences are to derive meaning from assessment results. Whether in relation to expected performance based on professional judgment, or in relation to a normative standard, comparative information should be as much a part of the assessment report as the raw data.

Interpretive information about the implications of the assessment results are necessary to ensure that the information will be used. In his monograph, Frank Womer urged that interpretation be built into the total plan. Interpretation, as to evident strengths and weaknesses and possible courses of action, can only be made by subject matter experts; but it is the responsibility of the assessment staff to collect and report the interpretive information as well as the pupil performance data. National Assessment arranges for subject matter

experts to react to the assessment results and to recommend public policy changes that may be warranted at the national level or curricular changes that might be desirable at the local level. Oregon and Maine are two of the few states where interpretive information has been provided as part of the reporting plan. Committees of teachers and reading specialists outlined what results could mean and what possible responses would be appropriate from state and local agencies.

12. Field Test Reports.

This final recommendation does not occur among the old rules for successful reporting. Field testing is the best means for implementing the advice offered in the foregoing principles. There is no way of anticipating the exact content, vocabulary, or organization that will be best for a particular audience. Reports should be tried out with their intended users in the same way that assessment instruments are field tested. Initially, a small number of respondents is required. Once major flaws in wording or formating have been eliminated, more systematic sampling should be done to learn from users what type of report will be most effective. Field testing should be done long before actual assessment results are available, either by using old test scores to approximate the new assessment data or by simulating results. In this way, an improved reporting format will be available by the time that the results are analyzed.

Follow-up studies should be used to facilitate the continued improvement of assessment reports. Interviews with a selected sample from an intended audience will provide feedback as to the actual meaning inferred from the narrative and data displays in a report. More extensive sampling can also be done by questionnaire to learn of the actual uses made of assessment results. Unintended uses may be worth addressing directly in subsequent assessments. In both field testing and follow-ups, the principles enumerated in this paper should provide categories for asking audiences to evaluate the utility of a report.

Conclusion

The ultimate success of state assessment programs will depend on how well assessment results are reported to their various audiences. In this paper, the most compelling recommendations for improving reporting practices are principles one, two, and twelve: plan ahead, develop different reports for different audiences, and field test report formats to determine the language and content that are most meaningful to respective audiences. Reporting should receive the same careful attention as instrument construction with sufficient opportunity for feedback from intended users.

BIBLIOGRAPHY

Bettinghaus, E.P., & Miller, G.R. A Dissemination System for State Accountability Programs. Denver, Colo.: Cooperative Accountability Project, 1973.

Glass, G.V, & Stanley, J.C. Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice Hall, 1970.

Larsen, E.P. Opening Institutional Ledger-Books -- A Challenge to Educational Leadership. Suggestions for Talking to School-Community Groups about Testing and Test Results. TM Report 28. Princeton, N.J.: ERIC Clearinghouse on Tests, Measurement, & Evaluation, Educational Testing Service, 1974.

Womer, F.B. Developing a Large Scale Assessment Program. Denver, Colo.: Cooperative Accountability Project, 1973.